DOCUMENT RESUME

ED 204 578 CE 029 501

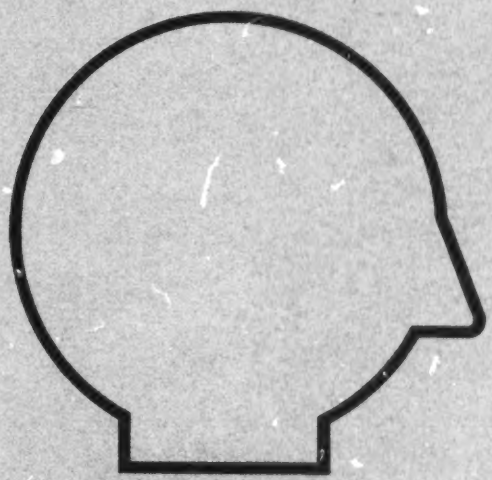
TITLE	Preparing Data for Analysis. Module 19. Vocational Education Training in Environmental Health Sciences.
INSTITUTION	Consumer Dynamics Inc., Rockville, Md.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, D.C.
PUB DATE	[81]
CONTRACT	300-80-0088
NOTE	49p.; For related documents see CE 029 482-507.
AVAILABLE FROM	National Technical Information Service, U.S. Dept. of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Charts; Competency Based Education; *Data Analysis; Educational Equipment; *Environmental Education; *Environmental Technicians; *Graphs; Learning Activities; Public Health; *Statistical Data; Tests; Vocational Education
IDENTIFIERS	Environmental Health

ABSTRACT

This module, one of 25 on vocational education training for careers in environmental health occupations, contains self-instructional materials on preparing data for analysis. Following guidelines for students and instructors and an introduction that explains what the student will learn are three lessons: (1) constructing a table by organizing data into rows and columns; (2) constructing graphs, including a line graph, a histogram, and a frequency polygon, by displaying the relationship between two sets of exact data; and (3) constructing charts, including bar charts and geographic coordinate charts, by displaying statistical information using only one set of exact data. Each lesson contains objectives, recommended methods and locations for practice, performance criteria, equipment and supplies to perform a task, detailed step-by-step instructions for learning a task, and performance exercises. Four performance tests cover tabulating data for analysis; preparing graphs; preparing histograms and frequency polygons; and preparing bar charts and maps. (CT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204578



Preparing Data for Analysis

Module 19

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CE 029 501

U.S. DEPARTMENT OF EDUCATION, Office of Vocational and Adult Education

FOREWORD

The Curriculum and Instruction Branch of the Office of Vocational and Adult Education, U.S. Department of Education, identified a need to improve the training opportunities for vocational education students interested in pursuing careers in environmental health. To fulfill that need, Consumer Dynamics, Inc., a Rockville, Maryland, based company, was awarded the contract to develop performance-oriented, competency-based modules in the environmental health sciences.

PREPARING DATA FOR ANALYSIS is one of the modules in the series, "Vocational Education Training in Environmental Health Sciences." The module content is based on selected materials in the environmental health field. The module is intended to supplement existing course materials.

CONTENTS

FOREWORD i

USING THESE SELF-INSTRUCTION MATERIALS 1
- Guidelines For Students 1
- Guidelines For Instructors 2

INTRODUCTION 4
- Background 4
- What You Will Learn 6

LESSON ONE 7
- Objective 7
- Where And How To Practice 7
- How Well You Must Do 7
- Things You Need 7
- Getting There--Steps 8
- Exercises 12

LESSON TWO 15
- Objective 15
- Where And How To Practice 15
- How Well You Must Do 15
- Things You Need 15
- Getting There--Steps 21
- Exercises 27

LESSON THREE 29
- Objective 29
- Where And How To Practice 29
- How Well You Must Do 29
- Things You Need 29
- Getting There--Steps 35
- Exercises 40

PERFORMANCE TEST 41
- Tabulating Data For Analysis 41
- Preparing Graphs 42
- Preparing Histograms and Frequency Polygons 40
- Preparing Bar Charts And Maps 43

REFERENCES 45

USING THESE SELF-INSTRUCTION MATERIALS

This self-instruction learning package or module is designed to allow both students and instructors flexibility of use. Although primarily intended for use in existing training programs, the module can be used by anyone interested in learning new skills or perking up old ones. Therefore, two sets of guidelines are presented--one set addressed to students and the other set addressed to instructors. First, find out how you, the student, should use the materials in this book.

GUIDELINES FOR STUDENTS

Take the Performance Test as a pretest.

When you pick up this book and work through it, your goal will not be a letter grade or a high score on an exam. Instead, you will work to develop skills that you can measure. You will not have to worry about how well someone else is doing. Before you start work on this module, you should, first, find out if you have sufficient skills to start training by reading through the section called PERFORMANCE TEST. If you think you can do all or most of the items in this test, ask your instructor to obtain the necessary data and supplies. You should have had a course in high school mathematics and have a basic knowledge of data collection and preparation techniques.

Work on parts you need to practice.

If you do everything well, according to the criteria in the Performance Test guidelines, you will not need to spend time working on this module. If after taking the Performance Test you discover there are parts you need to practice, follow the key to each item in FOR FURTHER STUDY.

Work straight through each lesson in the order presented.

Should you decide to completely work through this module, begin with the INTRODUCTION and go straight through each of the three lessons. The lesson begins with the OBJECTIVE of the training. Follow the instruction for each part in the order presented. Practice each step in a lesson until you can do it according to the criteria stated for the step. At the end of a lesson, do the EXERCISES. When there are audiovisuals listed at the end of a lesson, ask your instructor for help in obtaining them.

Take the Performance Test as a posttest.

Finally, after you have mastered all of the exercises in each lesson, ask your instructor to watch you do each item in the Performance Test. The items in the Performance Test are intended for use as a posttest to evaluate the quality of your performance. Turn now to the Performance Test.

GUIDELINES FOR INSTRUCTORS

Approach

The approach of these materials is to provide the student with the opportunity to learn skills of preparing data for analysis. In several steps you will have to supply data or arrange for students to collect their own data. These data should contain three or four interrelated variables. At least one of the variables should have subvariable components. Supply subvariables if possible.

Independent Study

Students can work independently and at their own pace. Depending on the time frame you set for completing each lesson, you may want to start a group off in each lesson with a demonstration and informal presentation.

As a Laboratory Workbook

Alternatively, you may choose to use this module as a laboratory workbook in a structured laboratory session. With this option, you may allow students greater access to your assistance, especially in watching them perform the pre- and posttest portions of the training.

General Instructions

Read through each lesson to anticipate what data and supplies you will need to make available for students to use. Also, order any audiovisuals or reading materials you think may present a complementary perspective to the training in this module. Use the items in the Performance Test as the minimum requirements for gauging successful completion of the training.

Specific Instructions

In this module, students are asked to prepare data for analysis using a number of basic formats: tables, graphs, and charts. Have on hand a number of reference materials that supply data for student use.

In Lesson One you should provide raw (untabulated) data or arrange for students to collect their own. Be sure the data provided has interrelated variables and that at least one of the variables has subdivisions. If possible, provide one or more variables with sub-subdivisions. Two sets of data are recommended, one set for the steps (see Step 1) and another set for the exercises (see Instruction 3).

More data should be provided for the Performance Test. Under "Tabulating Data For Analysis," you should supply a set of raw data. Under "Preparing Graphs," you should supply tabulated data (you may select data from a Key Point in Lesson One). Under "Preparing Histograms and Frequency Polygons," you should supply more tabulated data (you may select data from another Key Point in Lesson One). Under "Preparing Bar Charts and Maps," you should again supply tabulated data (here, also, you may use data from a Key Point in Lesson One).

INTRODUCTION

BACKGROUND

Numerical information (data) must be organized to make it meaningful. Imagine your checkbook with random, unnumbered, or undated entries: The addition of "organizers" such as serial ordering and dates makes life easier and information more useful. The need for organized information is critical to science and technology so that data may be efficiently and effectively analyzed and communicated to provide answers to questions and to provide bases for decision-making. Since organized data are essential to successful science and technology programs, a variety of ways to organize data have been developed.

Numerical information deals with variables: each property that can appear in different objects in distinctly different values. Typical variables are mass, number, size, speed, color, age, temperature, force, and so on. Most variables can be further divided into subvariables and sub-subvariables. For example, red and blue are subvariables of color. Usually, numerical information is first organized by tabulating it. A table is a systematic arrangement of data in columns and rows. Each variable, subvariable, or sub-subvariable normally occupies a single column, with its data running down the page; rows run across the page and can show relationships between variables at separate intervals. A table is used to present the data that can be easily read; therefore, a table should be as simple as possible. Two or three small tables usually are preferred to a single large table containing many details or variables. Generally, three variables are the maximum number that can be read with ease.

It is from tables that graphs and charts are prepared. Patterns and trends in data are difficult to identify and interpret in tables when there is a large amount of information. Graphs and charts help solve this problem by making patterns and trends more visible.

A graph is a method of showing quantitative data using a system of two or three coordinates. Typically, each coordinate deals with one variable. Among the kinds of graphs usually employed by environmental health professionals are line graphs, histograms, and frequency polygons. A line (or curve) graph shows variations in data by plotting points whose positions are determined by their respective values in two variables. A histogram is a bar graph of the number of times (frequency) an event occurs in subdivisions of a specific variable. A frequency polygon shows the frequency of events from two or more sets of data (variables) in a single diagram.

INTRODUCTION/BACKGROUND

Charts are methods of illustrating numerical information using only one coordinate. They are especially appropriate for comparing the magnitudes of different variables or of components of a whole. Two kinds of charts generally are used by environmental health professionals: bar charts and map charts. Bar charts are used to show and compare magnitudes of events. They can also be used to show frequency distributions and time series data. Map charts are used to show the occurrence of events. Spot maps use dots or other symbols to show locations where events occurred or where a specific condition exists. Area maps use shaded or coded zones to show the frequency of an event in subareas or the distribution of some condition. All of these data-organizing techniques are intended to make information collected by environmental health professionals more meaningful.

WHAT YOU WILL LEARN

When you finish working through the steps and exercises in this module, you will be able to construct a table, a graph, and a chart, using any of a wide variety of environmental health data.

Figure 1. A General Bar Chart

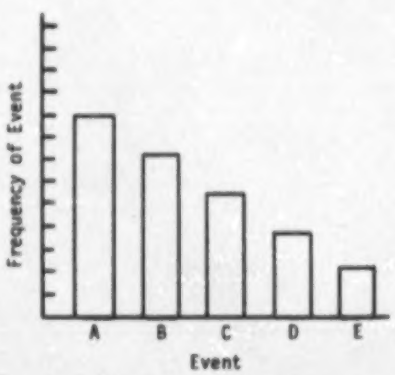
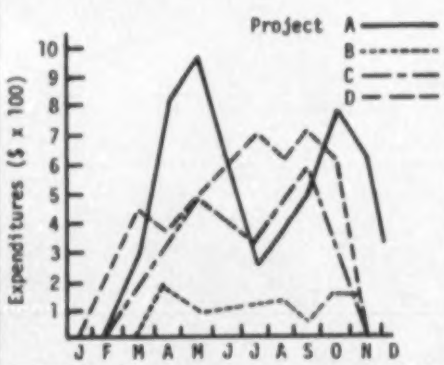


Figure 11. Monthly Expenditures on Four Projects, 1979.



You will learn how to organize data for analysis or for presentation to others in these three lessons:

- o Lesson One

  You will be able to construct a table by organizing data into rows and columns.

- o Lesson Two

  You will be able to construct graphs, including a line graph, a histogram, and a frequency polygon, by displaying the relationship between two sets of exact data--one set lying along the horizontal axis and the other along the vertical axis.

- o Lesson Three

  You will be able to construct charts, including bar charts and geographic coordinate charts, by displaying statistical information using only one set of exact data.

LESSON ONE

OBJECTIVE

You will be able to construct a table by organizing data into rows and columns.

WHERE AND HOW TO PRACTICE

This lesson can be performed anywhere you can find a clean table-top. Read through each step before working on it. If you have any questions about anything in the lesson, ask your instructor or supervisor for help. When you think you know how to construct a table by organizing data into rows and columns, practice this skill in the exercises.

HOW WELL YOU MUST DO

You must be able to correctly complete each instruction in the exercises.

THINGS YOU NEED

You will need the following materials to complete this lesson:

- paper
- pencil
- ruler
- two sets of data, each with three or four variables. (Ask your instructor to supply the data or to check those you supply yourself.)

Instructions: Now turn to the next page and begin work on Lesson One, "Getting There--Steps."

GETTING THERE--STEPS

STEP 1

Numerical information (data) is organized by putting it into a table. Most numerical information consists of several related variables, or factors that can change or have a number of different values. Common variables may be factors such as time, place, temperature, volume, concentration, speed, number, and weight. Obtain a set of data from your instructor that contains three or four variables. You will use these data in the following steps.

KEY POINT 1

A variable is a quantity or characteristic that can be measured or observed.

STEP 2

To construct a table, you must decide which of the data belongs to one variable, which belongs to a second variable, a third variable, and so on. Divide the data you obtained in Step 1 into the separate variables. List the numerical information for each variable in separate columns.

KEY POINT 2

Height	Weight	Age
75 in	190 lb	19 yr
62 in	148 lb	15 yr
68 in	172 lb	29 yr
66 in	130 lb	16 yr

The numerical information for each variable (e.g., weight) is listed in a separate column.

STEP 3

A table's data are displayed in rows and columns: Rows read across the page; columns read down the page. The left-most column, or stub, identifies entries on the rows. The stub generally contains entries that tell what the other variables apply to. The dates in the Key Point 3 table are the stub column. They tell when different flow rates were recorded. Choose a stub column from your lists in Step 2. Record your stub column on the left side of a page and list one of the other variables to the right.

STEP 4

Tables should be self-explanatory. Labels are important in easy-to-read tables. Each column is labeled at the top to identify the variable. A unit of measure for each variable is shown in abbreviated form above each column. Add another variable to your Step 3 list so that you have three columns. Label each column and show its unit of measure. Box the labels as shown in Key Point 4.

KEY POINT 3

Date (1980)	Flow (l/s)
Jul 6	1550
Jul 12	1200
Jul 22	1410
Jul 30	1530
Aug 7	1650
Aug 15	1700
Aug 21	1570
Aug 30	1320
Sep 3	1100
Sep 11	1140
Sep 17	1210

In this table the date is the stub column variable. Each row shows the rate of flow measured on a given date.

KEY POINT 4

Date (1980)	Flow (l/s)	Suspended Solids (mg/l)
Jul 6	1550	49
Jul 12	1200	53
Jul 30	1410	42
Aug 7	1530	39
Aug 15	1650	41
Aug 21	1700	52
Aug 30	1570	39
Sep 3	1320	61
Sep 11	1140	42
Sep 17	1210	46

Labels and units of measure are given for each variable. Here, (l/s) refers to a flow rate of liters per second, and (mg/l) refers to the concentration of suspended solids in milligrams per liter.

STEP 5

Data from a single variable are sometimes divided into subvariables. Just as each variable has its own label, each subvariable must be labeled. The table in Key Point 5 shows two major variables: age and number of cases (disease incidence), which is divided into two labeled subvariables according to race. The subdivided variable is never placed in the stub column. A straddle line separates the subvariable headings from the major variable heading. Following Steps 1-4, list a variable with several subvariables, label each, and include a straddle line. Vertical lines separating the columns are optional.

KEY POINT 5

Age Group (yrs)	No. of Cases by Race	
	White	Other
0-14	674	568
5-14	526	555
15-24	1,263	1,219
25-44	4,017	4,136
44-64	6,841	4,015
65+	5,171	1,853

STEP 6

A variable that has been subdivided once may be sub-subdivided. Each level of the variable, the major variable, the subdivisions, and the sub-subdivisions must be labeled and separated by straddle lines. The example in Key Point 6 shows a variable (number of cases) that has been subdivided by race and sub-subdivided by sex. Following Steps 1-5, select data that can be subdivided and sub-subdivided. List and label all divisions and add straddle lines Box all parts.

KEY POINT 6

Age Group (yrs)	No. of Cases by Race and Sex			
	White		Other	
	Male	Female	Male	Female
0-14	620	580	593	530
5-44	3,126	2,154	3,194	2,161
45-64	5,198	1,643	3,008	1,067
65+	3,556	1,715	1,235	618

STEP 7

A title identifies the data in a table. It describes the content as to the subject, place, and time. The title should be numbered. In Key Point 7, the title answers the questions of what, when, and where the data were collected. The source of the data is identified below the table. Using the table you prepared in Step 6, write a correct title indicating what the subject is and when and where the data were recorded. Indicate your source of data below the table.

KEY POINT 7

TABLE 1. NO. OF NEWLY DIAGNOSED CASES OF TUBERCULOSIS BY AGE, RACE, AND SEX, UNITED STATES, 1973*

Age Group (yrs)	No. of Cases by Race and Sex			
	White		Other	
	Male	Female	Male	Female
0-14	620	580	593	530
15-44	3,126	2,154	3,194	2,161
45-64	5,198	1,643	3,008	1,067
65+	3,556	1,715	1,235	618

*Source: HEW, PHS, CDC, Reported Morbidity and Mortality, 1974. Vol. 23, No. 53, p 12.

The title shows what the subject is and where and when the data were collected. The source of the data is indicated below the table.

EXERCISES

Instruction 1: The following tables have several variables in columns that appear in random order. Select a column for each table that would be the best stub column and explain your answer to your instructor. Circle your answer.

TABLE 1. MOVING AVERAGE AND RANGE

Column a Moving Average	Column b Assay Value	Column c Sample No.	Column d Moving Range
17.2	17.1	1	+0.3
17.4	17.3	2	+0.0
17.3	17.4	3	-0.1
17.1	17.2	4	-0.2
16.9	17.0	5	-0.1

TABLE 2. RELATIVE RESPONSE OF SOUND LEVEL METER WEIGHTING NETWORKS

Column a Weighted Response (dB)			Column b Frequency (Hz)
A	B	C	
-39	-17	-3	31.5
-26	- 9	-1	63.0
-16	- 4	-0	125.0
- 9	- 1	-0	250.0
- 3	- 0	-0	500.0

TABLE 3. DOSE BUILDUP FACTOR FOR A POINT ISOTROPIC SOURCE

Column a μX			Column b	Column c
1	2	4	MeV	Material
3.09	7.14	23.0	0.25	Water
2.52	5.14	14.3	0.5	
2.37	4.24	9.47	0.5	Aluminum
2.02	3.31	6.57	1.0	
1.98	3.09	5.98	0.5	Iron

Instruction 2: The following tables have variables, subvariables, and sub-subvariables that need to be arranged; straddle lines and measurement units must also be added in the boxes provided. Each box has the stub column already labeled. The other labels for variables, subvariables, and units are randomly listed below the box. Confirm your arrangement with your instructor.

TABLE 4. POPULATION (BY AGE, RACE, AND SEX), THATTOWN, USA, 1975

Age Group (yrs)

Female

Black

Population

Male

White

Male

Female

TABLE 5. WEATHER CHARACTERISTICS, THISTOWN, USA, 1975-80

Year (Stub Column)

mph

Rainfall

Day

Wind

Month

Temperature

Inches

High

Degrees (F)

Low

Instruction 3: Repeat Lesson One using another set of data supplied by your instructor. Choose a stub column variable, divide variables and subvariables, add measurement units, and title the table. Check your table with your instructor. When you can correctly tabulate a set of data, you are ready to being work on Lesson Two.

LESSON TWO

OBJECTIVE

You will be able to construct graphs, including a line graph, a histogram, and a frequency polygon, by displaying the relationship between two sets of exact data--one set lying along the horizontal axis and the other along the vertical axis.

WHERE AND HOW TO PRACTICE

This lesson can be performed anywhere you can find a clean table-top. Read through each step before working on it. If you have any questions about anything in the lesson, ask your instructor or supervisor for help. When you think you know how to construct line graphs, histograms and frequency polygons, practice these skills in the exercises.

HOW WELL YOU MUST DO

You must be able to complete correctly each instruction in the exercises.

THINGS YOU NEED

You will need the following materials to complete this lesson:

- graph paper (for example, use graph paper that is gridded into 10 x 10 squares per inch with every 10th line heavy)
- pencil
- ruler
- Data Sets A-G (pages 16-20 of this module).

DATA SET A

Average Pulse Rate for Patient #1 from 7 a.m. to 7 p.m.

Time	Pulse Rate (beats/min)
7 a.m.	65
8	65
9	67
10	70
11	71
12 p.m.	69
1	70
2	68
3	67
4	68
5	68
6	67
7	68

Source: Invented

DATA SET B

Particle-Size Distribution for Aerosol Type X

Size (μm)	Number of Particles/cm^3
0.5	10
0.7	22
1.0	26
1.4	29
2.0	37
2.7	28
3.8	22
5.4	14
7.7	8
10.9	4

Source: Invented

DATA SET C

Average Pulse Rate for Patient #2 from 7 a.m. to 7 p.m.

Time	Pulse Rate (beats/min)
7 a.m.	68
8	69
9	69
10	71
11	72
12 p.m.	74
1	72
2	71
3	71
4	69
5	68
6	69
7	68

Source: invented

DATA SET D

Particle-Size Distribution for Aerosol Type Y

Size (μm)	Number of Particles/cm^3
0.5	6
0.8	13
1.2	21
1.6	24
2.0	30
2.5	37
2.9	45
3.5	39
5.6	22
8.0	16
10.5	7

Source: invented

DATA SET E

Number of Reported Cases of Influenza, Sample City, 1970

Week of Onset of Illness	No. of Cases
1	0
2	6
3	7
4	12
5	15
6	10
7	5
8	2
9	0

Source: invented

DATA SET F

Number of Reported Cases of Influenza Types A, B, and C by Week of Onset, Sample City, 1970

Week of Onset of Illness	No. of Influenza Cases			
	Type A	Type B	Type C	Total
1	0	0	0	0
2	4	0	2	6
3	4	1	2	7
4	8	1	3	12
5	10	2	3	15
6	6	2	2	10
7	3	1	1	5
8	1	1	0	2
9	0	0	0	0

Source: invented

DATA SET G

Expenditure on Laboratory Equipment
for Project X, 1979-80

Month	Expenditure ($)
Dec	0
Jan	50
Feb	80
Mar	150
Apr	0
May	100
Jun	50
Jul	30
Aug	30
Sep	20
Oct	0
Nov	10
Dec	0

Source: invented

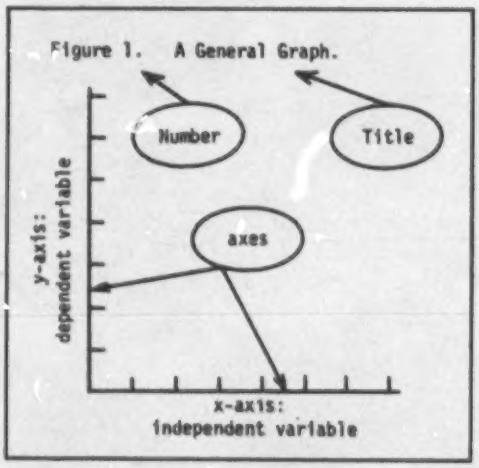
Instructions: Now turn to the next page and begin work on Lesson Two, "Getting There--Steps."

GETTING THERE--STEPS

STEP 1

To draw a graph, you must first choose the axes (right angle lines) on which to place the variables from your data. Usually the variable on the horizontal (x) axis is the independent variable (the method of classification, such as time or age), while the variable on the vertical (y) axis is the dependent variable (often the frequency of occurrence of an event). Find Data Set A in "Things You Need." Identify the independent and dependent variables. On graph paper draw a set of axes; place (label) the Set A variables on the proper axis. Number and title your graph.

KEY POINT 1

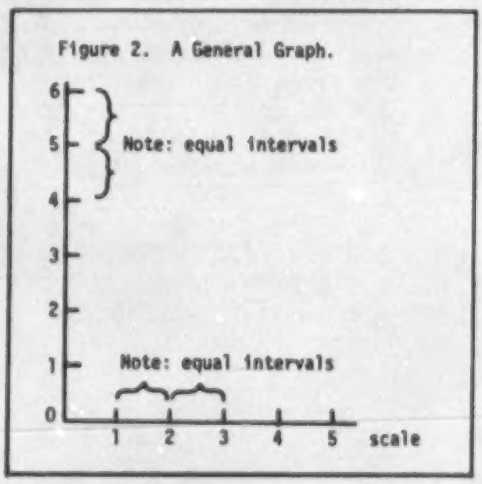


The vertical (y) axis shows the dependent variable; the horizontal (x) axis shows the independent variable.

STEP 2

Choose a scale (progressive graduated series) for each axis in Step 1. The axis may have different scales, but on each the intervals must represent equal numerical values. Identify the range of the variables in Set A. Count the number of spaces on each axis needed to fit the data. Extend the axis slightly beyond the ranges. A wide range of numbers can be displayed in a small space by proper scale selection. Label the scale divisions and specify the units (e.g., hour) into which the scale is divided.

KEY POINT 2



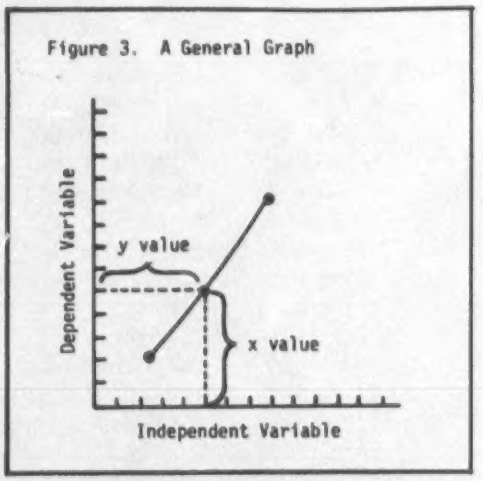
Equal interval scales make up each axis. The axes and the units in each scale are labeled.

STEP 3

Locate and mark the data points from Set A on the graph prepared in Step 2. To find each point, first locate the value of an independent variable on the x-axis. Move your finger or a pencil vertically upward to the position that is opposite the corresponding value of the dependent variable on the y-axis. Mark this point. Continue to mark points for all of the data in Set A. Connect the points with a line.

KEY POINT 3

Figure 3. A General Graph



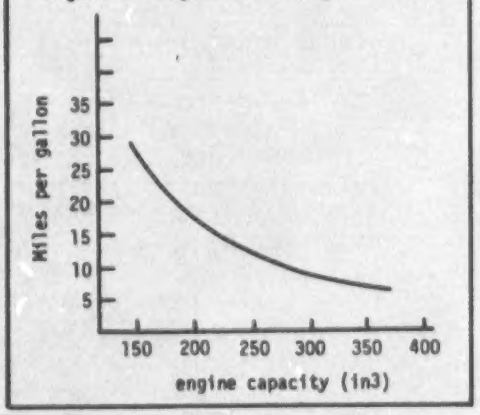
Each point on the graph shows where the values of an independent and a dependent variable meet.

STEP 4

Repeat Steps 1, 2, and 3 using Data Set B. These data illustrate how changes in one factor, the independent variable (particle size), affect another factor, the dependent variable. Draw axes, choose the proper axis and scale for each variable, label the axes and scales, plot and connect the data points, and number and title the graph.

KEY POINT 4

Figure 4. Engine Efficiency.

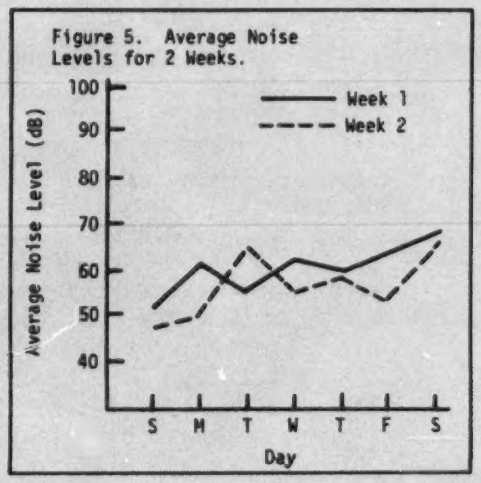


Plotted data show the relationship between dependent and independent variables.

STEP 5

Graphs can be used to show more than one set of data. Key Point 5 shows how two data sets (average noise levels for 2 weeks) can be plotted on the same axis in a single figure. (Note: A different kind of line, identified by a legend, is used for each set of data.) Following Key Point 5, repeat Steps 1, 2, and 3, plotting two lines from Data Sets A and C on the same graph. Then repeat Step 4 plotting two lines from Data Sets B and D.

KEY POINT 5

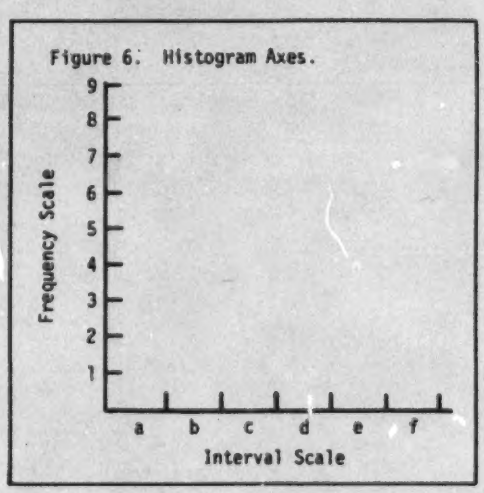


Two data sets can be shown on the same figure. Each data set is represented by a different line and identified in a legend.

STEP 6

A histogram uses data from a frequency distribution (the number of times an event occurs in given intervals). In histograms, the interval scale usually appears on the horizontal axis and the frequency scale appears on the vertical axis. The scale for each axis is usually continuous. Draw and label a set of axes for Data Set E.

KEY POINT 6



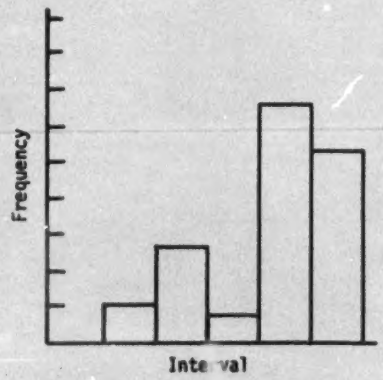
In a histogram, the x-axis usually shows the event intervals and the y-axis the frequency of events.

STEP 7

Select a scale for each axis in Step 6. As in graphing (see Step 2), all scale units on each axis are identical in size and are continuous. Choose a scale for each variable in Data Set E and add them to the axes. Label each axis and scale. Add vertical spaces of the same width to show the frequency of occurrence of the event in each interval. Give the histogram a complete title, including number, identity of data, place, and time.

KEY POINT 7

Figure 7. A General Histogram.



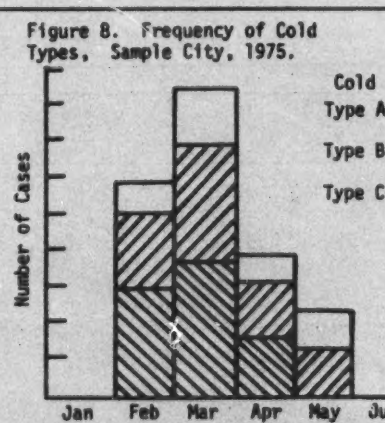
Scales on each axis are consistent. Vertical bars show frequency of event.

STEP 8

The bars of a histogram may show the proportionality of a frequency distribution, or show how several components combine to make a total. Variable components (subtotals) are stacked on each other, each with its own identity, to make the total. A legend identifies each component. Using Key Point 8 as a model, and Data Set F, draw a histogram that includes subtotals in each bar. Label the axes and scales, provide a legend, and give the histogram a number and title.

KEY POINT 8

Figure 8. Frequency of Cold Types, Sample City, 1975.



Show how subtotals make up the total by adding coded bars.

STEP 9

Frequency polygons, like histograms, use frequency distribution data. Instead of bars, a frequency polygon is constructed by plotting individual values of the event at the midpoints of their respective intervals and connecting them with a line. Key Point 9 shows the histogram from Step 7 displayed as a frequency polygon. (Note: The histogram is included only to show that the two forms represent the same data; do not show the corresponding histogram on the same graph with a frequency polygon.)

KEY POINT 9

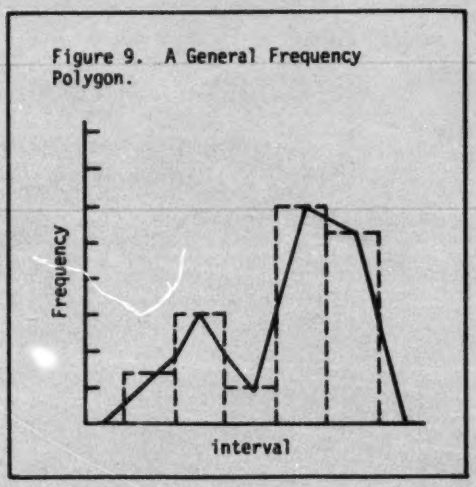


Figure 9. A General Frequency Polygon.

Plot the data points (including zero) at the interval midpoints. Do not put a histogram in the same figure.

STEP 10

Using Data Set G, and following Steps 7-9, set up axes and scales and plot a frequency polygon. When the frequency of an event in a given interval is zero, the data point falls on the horizontal axis at the midpoint of that interval as shown in Key Point 9. When you have completed the frequency polygon for Data Set G, check to see that you have labeled each axis and scale and have given the figure a number and title.

KEY POINT 10

In frequency polygons, zero points are always marked on the x-axis. Titles and labels are always given.

STEP 11

Frequency polygons often present more than two sets of data. Following the example in Key Point 11, use Data Set F to construct a frequency polygon that includes data from four sub-variables. Use four kinds of lines and include a legend to indicate what each line represents.

KEY-POINT 11

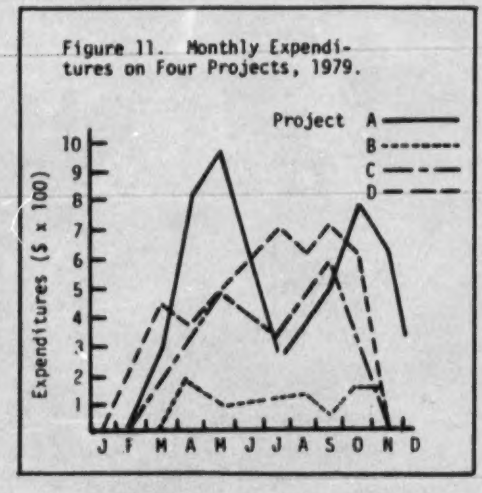


Figure 11. Monthly Expenditures on Four Projects, 1979.

A legend is used when more than one set of data are shown in a frequency polygon.

EXERCISES

Instruction 1: Graph the following data in a single figure. Be sure to complete the figure by adding appropriate labels, numbers, and title. Check the finished graph with your instructor.

TABLE 1. MONTHLY AVERAGE RATE OF FLOW OF THE LONG RIVER, 1978-80

Month	Average Flow (l/m)		
	1978	1979	1980
Jan	1,960	2,038	1,996
Feb	2,050	2,090	1,961
Mar	2,435	2,384	2,111
Apr	2,671	2,609	2,452
May	2,390	2,462	2,582
Jun	2,010	2,211	2,413
Jul	1,842	1,939	2,207
Aug	1,813	1,902	2,102
Sep	1,782	1,857	2,183
Oct	1,886	1,879	2,067
Nov	1,945	1,922	2,008
Dec	2,003	1,968	1,972

Instruction 2: Plot a histogram of the following data in a single figure. Label and title the figure. Check the finished histogram with your instructor.

TABLE 2. FISH CAUGHT IN ROUND LAKE OVER MANY YEARS

Date	Number of Fish Caught			
	Perch	Cisco	Carp	Total
1910-1919	2	26	2	30
1920-1929	12	20	6	38
1930-1939	24	9	10	43
1940-1949	9	0	5	14
1950-1959	0	0	0	0
1960-1969	12	0	5	17
1970-1979	14	2	9	25

Instruction 3: Using the data in Table 2, plot a frequency polygon. Check the finished frequency polygon with your instructor. When you can correctly produce graphs, histograms, and frequency polygons, you are ready to being work on Lesson Three.

LESSON THREE

OBJECTIVE

You will be able to construct charts, including bar charts and geographic coordinate map charts, by displaying statistical information using only one set of exact data.

WHERE AND HOW TO PRACTICE

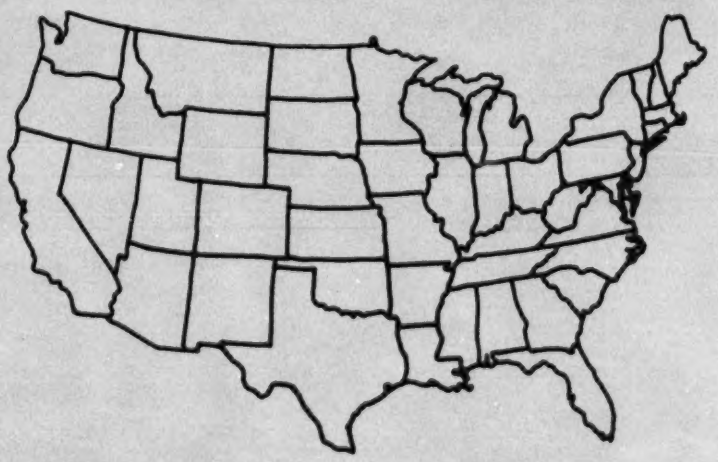
This lesson can be performed anywhere you can find a clean tabletop. Read through each step before working on it. If you have any questions about anything in the lesson, ask your instructor or supervisor for help. When you think you know how to construct bar charts and geographic coordinate map charts, practice these skills in the exercises.

HOW WELL YOU MUST DO

You must be able to complete correctly each instruction in the exercises.

THINGS YOU NEED

- graph paper
- pencil
- ruler
- map of the United States showing locations of States and principal cities
- six copies of the following outline map of the United States:



THINGS YOU NEED (cont'd)

o Data Sets A-F (pp. 30-34 of this module).

DATA SET A

REPORTED CASES OF RABIES IN HUMANS BY YEAR,
UNITED STATES, 1950-60

Year	Number of Reported Cases
1950	16
1952	24
1954	14
1956	10
1958	5
1960	2

DATA SET B

PERCENTAGE OF MALES AND FEMALES IN VARIOUS AGE GROUPS,
ANY CITY, 1970

Age Group	% of Population	
	Male	Female
1	46	54
1-4	49	51
5-14	50	50
15-44	48	52
45+	45	55

DATA SET C

POLLUTANTS ENTERING
BEA POND, 1970-79

Year	Pollutants (kg)		
	Inorganic	Organic	Total
1970	4	28	32
1971	6	27	33
1972	14	23	37
1973	11	24	35
1974	22	22	44
1975	23	26	49
1976	24	28	52
1977	20	25	45
1978	20	21	41
1979	17	19	36

DATA SET D

POPULATION DENSITY

City	Persons/sq mi
Albuquerque	205
Atlanta	492
Austin	256
Baltimore	1082
Birmingham	404
Boston	940
Buffalo	548
Chicago	1039
Cincinnati	655
Cleveland	788
Columbus	600
Dayton	598
Denver	455
Detroit	893
Fort Worth-Dallas	382
Houston	338
Indianapolis	593
Jacksonville	317
Kansas City	420
Los Angeles	516
Louisville	555
Memphis	416
Miami	266
Minneapolis-St. Paul	500
New Orleans	454
New York	2872
Norfolk	582
Oklahoma City	324
Omaha	364
Philadelphia	1387
Pittsburgh	737
Portland	257
Richmond	621
Salt Lake City-Ogden	219
San Diego	252
San Francisco	692
Seattle	222
Spokane	162
St. Louis	872
Tampa-St. Petersburg	323
Washington, D.C.	1023
Wichita	308

DATA SET E

ACTIVE PROFESSIONAL NURSES (per 100,000 Civilians: 1962)

Rank	State	No.
1.	MA	502
2.	NH	491
3.	VT	445
4.	CT	440
5.	RI	396
6.	DE	391
7.	NY	388
8.	MN	384
9.	ME	371
10.	PA	371
11.	CO	370
12.	MT	349
13.	NJ	348
14.	OR	348
15.	ND	340
16.	WA	337
17.	AZ	335
18.	WI	332
19.	WY	325
20.	CA	324
21.	IA	320
22.	NE	320
23.	FL	302
24.	OH	295
25.	IL	291
26.	HI	288
27.	AK	287
28.	KS	284
29.	ID	276
30.	SD	270
31.	MI	266
32.	NV	262
33.	IN	248
34.	WV	248
35.	MD	246
36.	VA	235
37.	UT	234
38.	NC	231
39.	MO	220
40.	SC	214
41.	NM	213
42.	LA	198
43.	GA	194
44.	TN	177
45.	KY	175
46.	TX	172
47.	OK	163
48.	AL	157
49.	MS	142
50.	AR	120

DATA SET F

DEATH RATES BY ACCIDENT IN THE UNITED STATES, 1978

State	Death Rate (Per 100,000 population)	State	Death Rate (Per 100,000 population)
AL	66.1	MT	69.7
AK	60.6	NE	47.9
AZ	69.3	NV	73.4
AR	111.4	NH	38.2
CA	50.0	NJ	33.4
CO	51.6	NM	80.7
CT	33.9	NY	33.5
DE	41.4	NC	56.5
FL	50.2	ND	59.3
GA	58.6	OH	41.6
HA	35.5	OK	63.3
ID	69.3	OR	58.3
IL	42.9	PA	40.0
IN	47.6	RI	29.7
IA	47.8	SC	59.5
KA	54.0	SD	57.8
KY	52.6	TN	57.4
LA	59.9	TX	55.5
ME	46.1	UT	56.0
MD	38.1	VT	50.7
MA	40.0	VA	46.6
MI	43.3	WA	53.9
MN	48.5	WV	57.2
MS	72.3	WI	42.6
MO	53.6	WY	79.8

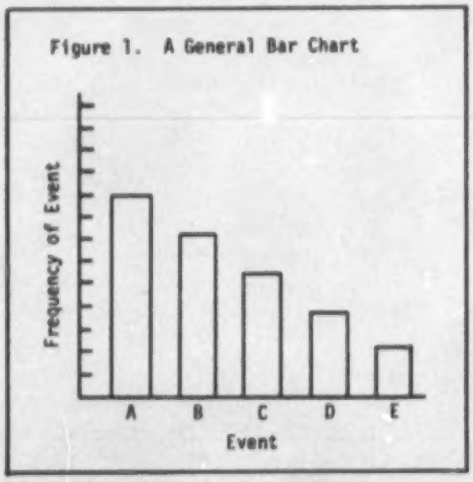
Instructions: Now turn to the next page and begin work on Lesson Three, "Getting There--Steps."

GETTING THERE--STEPS

STEP 1

Bar charts show comparative magnitudes of different events or of components of an event. Unlike histograms, spaces separate the bars in bar charts. Depending on axis selection, bars may be placed either vertically or horizontally. An arbitrarily selected width is uniform for all bars. Bar length is proportional to the frequency of the event. Bars may be arranged randomly, serially (e.g., by time), or in ascending or descending order. Following Key Point 1, use Data Set A from "Things You Need" to draw a bar chart. Number and title the figure and label each axis and scale.

KEY POINT 1

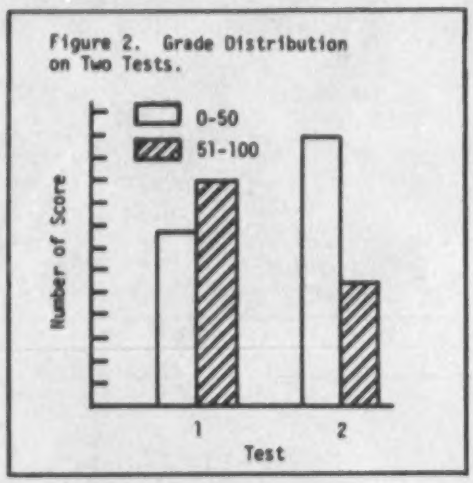


Bar charts show comparative magnitudes of different events.

STEP 2

Each interval in a bar chart may contain two or more bars to show comparative proportions. If an interval has more than one bar, each should be illustrated distinctively by a code. The codes must be shown in a legend, as in Key Point 2. Using Data Set B, draw and label a bar chart showing comparative sizes within a population.

KEY POINT 2

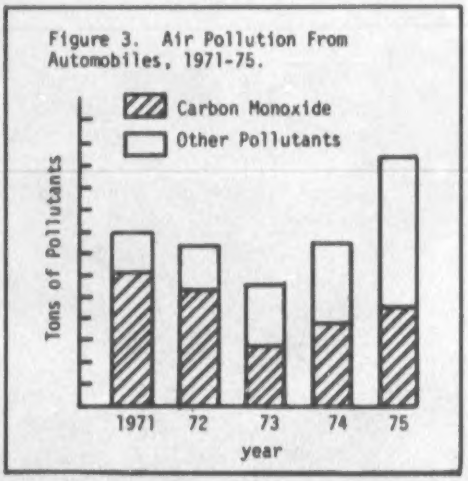


Two or more bars in a single interval show comparative proportions of event components.

STEP 3

Bar charts may also be used to show comparative proportions of a total and the frequency distribution of an event over time. Key Point 3 has a bar chart that illustrates both uses. The total bar height represents the total frequency of the event. Bar components, each distinctively marked, show comparative subtotals. The events are displayed as they occurred over time. Following Key Point 3 and using Data Set C, complete a bar graph that shows components of an event and a frequency distribution over time.

KEY POINT 3

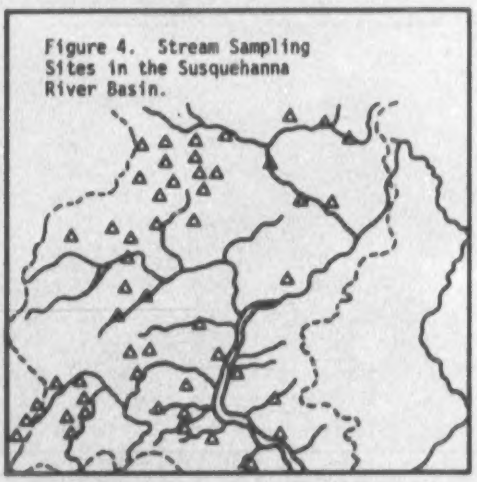


Bar charts can show frequency distributions over time.

STEP 4

Spot maps are made by placing a dot or other symbol on a map at the site where a situation exists or an event occurred. The simplest spot maps use only one kind of symbol (e.g., a dot). Using an outline map of the United States, a map showing the locations of major U.S. cities, and the numbers from Data Set D, place a dot on the map at each location with a population density greater than 500 persons per square mile.

KEY POINT 4

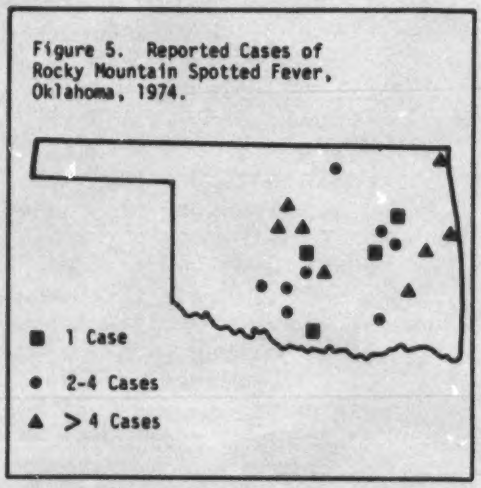


Spot maps portray the geographic distribution of events or conditions.

STEP 5

Spot maps can show geographic distributions of a variety of events or conditions. A legend identifies each symbol (see Key Point 5). Using an outline map of the United States, a map showing the location of principal U.S. cities, an atlas, and Data Set D, draw a spot map to show a variety of population densities. Determine the divisions by which you can group the data. For example, you may select three groups (<500, 501-1,000, >1,000). Assign a symbol to each group and place the appropriate symbol at the proper map locations. Include a legend and map title.

KEY POINT 5

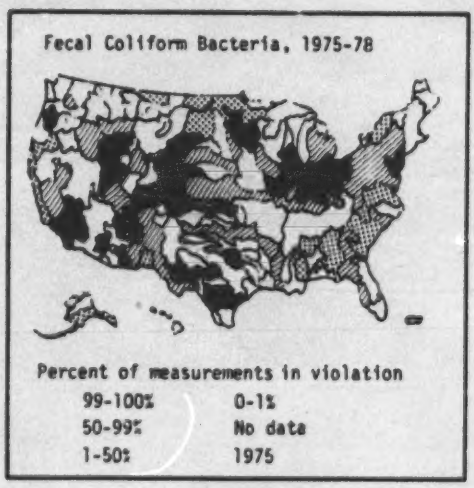


Symbols on spot maps show various frequencies or concentrations.

STEP 6

Area maps also show the geographic distribution of a condition or event. They use codes (visual symbols) to represent various distributions. Several methods can be used for grouping data. The simplest method is to arbitrarily divide the data at convenient breaks (see Key Point 6). Using an outline map of the United States, a map showing the locations of the States, and Data Set E, draw an area map showing the distribution of nurses. Choose at least four groups of numbers, select a code for each group, and complete the map with a legend and title.

KEY POINT 6



Codes in area maps show distributions of conditions or events.

STEP 7

Another method of selecting groups is to rank order (e.g., list from highest to lowest) all of the values to be coded. Using Data Set E, divide the list into five equal-sized and mapped. groups. Record the range of data (number of nurses) that correspond to the States that begin and end each group. Adjust the range of each group so there are no gaps between the end of one group and the beginning of the next, following this example:

KEY POINT 7

Area maps can show data grouped by rank. The entire data set is rank ordered and divided into equal-sized groups. The group ranges are adjusted to be continuous. Each group is coded

Group	Original Range of Numbers	Adjustment of Ranges	Adjusted Range
1	100-163	163 + 1/2 (175-163) = 167	100-167
2	175-198	198 + 1/2 (212-198) = 205	168-205
3	212-300		206-300

Assign an area code to each group and code each State using the code appropriate for that State's data. Complete the map with a legend and title.

STEP 8

The third method of grouping is to separate the data so that each group has the same size range. Using Data Set E, calculate the total range by subtracting the lowest data point from the highest. To calculate the range size for each group, divide the total range by the number of groups you choose. For example, to map two groups, divide the total range by two.

KEY POINT 8

Groups of equal range size can be shown. Determine the range of the data set and select the number of groups to be shown.

STEP 9

From Step 8, the total range is 382 (502-120=382). For two groups, each range is 191 (382 ÷ 2=191). Add the range size, 191, to the lowest data point, 120, twice to identify the range for each of your two groups:

A. 120 through (120 + 191)=
120 through 311

B. 312 through (120+191+191)=
312 through 502.

As in Steps 6 and 7, assign a code to each group and map the States. Repeat this procedure with four groups instead of two.

KEY POINT 9

Each group is assigned its own code. Codes are explained in a legend.

EXERCISES

Instruction 1: Using the data from Lesson One, Step 3, construct a bar chart. Label and title the figure. Check the bar chart with your instructor.

Instruction 2: Using the data for the first two age groups in Lesson One, Step 5, construct a bar chart with a single bar for each group showing the comparative proportions. Label and title the figure. Check the bar chart with your instructor.

Instruction 3: Develop a spot map of the words on this page showing, with different symbols, the position of all capital letters and punctuation. Label and title your figure.

Instruction 4: Using Data Set F and three outline maps of the United States, develop three different area maps as follows:

(1) A map in which groups are formed by arbitrary division.

(2) A map in which the groups have an equal number of States.

(3) A map in which each group has the same size range.

Check the area maps with your instructor.

PERFORMANCE TEST

Instructions: Check your skill level or progress by working through each of the items in this test. If you can perform each item as required, place a check in the space provided. When all of the items are checked, you are ready to demonstrate your skills to your instructor. You may use the following list if needed. You will be considered trained in a skill after your instructor approves your performance of each of the following items:

TABULATING DATA FOR ANALYSIS

No. 1 ____ Identify the variable, subvariables and sub-sub-variables, using a set of data provided by the instructor.

No. 2 ____ Select one variable for a table's stub column.

No. 3 ____ Set up and completely label the table.

No. 4 ____ List all data in the appropriate rows and columns.

No. 5 ____ Provide a complete title for the table.

FOR FURTHER STUDY

If you could not perform one or more of the five items above, review and practice the following lesson steps:

No. 1
Lesson One, Steps 1, 2, 5, and 6

No. 2
Lesson One, Step 3

No. 3
Lesson One, Steps 2 through 6

No. 4
Lesson One, Steps 2 through 6

No. 5
Lesson One, Step 7

PERFORMANCE TEST

PREPARING GRAPHS

No. 1 _____ Identify an independent and a dependent variable using data supplied by your instructor.

No. 2 _____ Select the proper axes on a graph for the independent and dependent variables.

No. 3 _____ Choose the appropriate scale for each axis of the line graph.

No. 4 _____ Label each axis correctly.

No. 5 _____ Plot data points accurately on the graph.

FOR FURTHER STUDY

If you could not perform one or more of the five items above, review and practice the following lesson steps:

No. 1
Lesson Two, Steps 1, 4, and 5

No. 2
Lesson Two, Step 1

No. 3
Lesson Two, Step 2

No. 4
Lesson Two, Steps 2 and 5

No. 5
Lesson Two, Steps 3, 4, and 5

PREPARING HISTOGRAMS AND FREQUENCY POLYGONS

No. 1 _____ Select the proper axes to construct a histogram using data supplied by your instructor.

No. 2 _____ Draw and label each histogram axis.

No. 3 _____ Select and add proper scales to the axes.

No. 4 _____ Complete the histogram by adding the appropriate bars.

PERFORMANCE TEST

No. 5 ____ Give the histogram a complete title.

No. 6 ____ Prepare a set of axes, with labels and scales, for a frequency polygon supplied by your instructor.

No. 7 ____ Complete the frequency polygon and include a title.

FOR FURTHER STUDY

If you could not perform one or more of the seven items above, review and practice the following lesson steps:

No. 1
Lesson Two, Step 6

No. 2
Lesson Two, Steps 6 and 8

No. 3
Lesson Two, Step 7

No. 4
Lesson Two, Steps 7 and 8

No. 5
Lesson Two, Step 7

No. 6
Lesson Two, Steps 9 and 10

No. 7
Lesson Two, Steps 10 and 11

PREPARING BAR CHARTS AND MAPS

No. 1 ____ Prepare a simple bar chart showing comparative frequencies using data supplied by your instructor.

No. 2 ____ Prepare a bar chart showing comparative proportions of frequencies.

No. 3 ____ Prepare a spot map using a map of the locations of principal U.S. cities, an outline map of the United States, and Data Set D, Lesson Three, showing locations with population densities of less than 200, 201-300, and 301-400 persons/square mile.

PERFORMANCE TEST

No. 4 ____ Prepare an area map using an outline map of the United States and Data Set E, Lesson Three, using the rank-order method and showing seven equal-sized groups.

No. 5 ____ Prepare an area map using an outline map of the United States and Data Set E, Lesson Three, showing six areas with the same size ranges.

FOR FURTHER STUDY

If you could not perform one or more of the five items above, review and practice the following lesson steps:

No. 1
Lesson Three, Step 1

No. 2
Lesson Three, Steps 2 and 3

No. 3
Lesson Three, Steps 4 and 5

No. 4
Lesson Three, Steps 6 and 7

No. 5
Lesson Three, Step 8

REFERENCES

U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control. Methods for Organizing Epidemiologic Data, Manual 4. Atlanta, GA, 1979.